

JAVITH NASEEM J

Data Scientist

Chennai | +919629272568 | javithnaseem.j@gmail.com | [Portfolio](#) | [linkedin.com/javithnaseemj](https://www.linkedin.com/in/javithnaseemj)

EDUCATION

Dr M.G.R Educational and Research Institute

B.TECH CSE (DATA SCIENCE & ARTIFICIAL INTELLIGENCE) [CGPA: 8.1]

Chennai, Tamil Nadu

Sep 2021- May 2025

Anjuman Matric Higher Secondary School

HIGH SCHOOL & HIGHER SECONDARY SCHOOL [Per: 83%]

Ayyampet, Tamil Nadu

Jun 2019 – May 2021

SKILLS

Languages: Python, SQ.

Libraries & Frameworks: Pandas, NumPy, Matplotlib, Seaborn, NLTK, Sci-kit Learn, TensorFlow, PyTorch,, FastAPI, Evidently-AI.

Tools: MS-Excel, PowerBI, Git, GitHub,Github-Action, MLflow, DVC, Docker, Kubeflow, Kubernetes, Prometheus, Grafana.

Platforms: AWS, PostgreSQL, MySQL.

Soft Skills: Communication, Problem Solving, Team Collaboration.

PROJECTS

Patients Condition Classifier and Drug Recommender : (GitHub)

Feb 2025 – Mar 2025

Tools & Technologies: [Python & libraries, DVC, Docker, GitHub Actions, Dagshub, Evidently AI]

- Built a multi-class classifier using TF-IDF to predict patient conditions from symptoms with over 90% accuracy.
- Implemented a recommendation engine that improved drug relevance by 32%, using user ratings and review text analytics.
- Enabled reproducibility using DVC for tracking data, models, and transformations within a modular pipeline setup.
- Tracked training runs on Dagshub and integrated GitHub Actions for CI/CD to automate workflow from commit to build.
- Incorporated Evidently AI to monitor drift and ensure real-time model performance in test production environments.

Bank Transaction Fraud Detection: (GitHub)

Mar 2025 – Apr 2025

Tools & Technologies: [Python & libraries, DVC, MLflow, Docker, Kubeflow, GitHub Actions, Observations, AWS]

- Developed a modular ML pipeline with DVC to automate ingestion, training, and evaluation of fraud detection models.
- Applied Optuna-based tuning for XGBoost, RF, and Logistic Regression; XGBoost achieved 94.7% F1 and 21% fewer false positives.
- Tracked over 100 experiments via MLflow auto logging and visualized results using ROC and confusion matrix dashboards.
- Packaged pipeline with Docker for reproducibility, enabling consistent model performance across environments.
- Versioned all data and artifacts using YAML configs and DVC, supporting seamless model rollback and iteration.
- Set up real-time drift detection and model health monitoring using Prometheus and Grafana in staging environment.

Predictive Analytics for Store Operations and Sales Optimization: (GitHub)

Apr 2025 – May 2025

Tools & Technologies: [Python & libraries, DVC, Docker, GitHub Actions, Observation, Evidently AI]

- Built a sales forecasting system using regression and time-series analysis, achieving a 28% reduction in RMSE.
- Designed custom features like promo periods, holidays, and weather effects to increase model interpretability.
- Versioned all pipeline stages, datasets, and results via DVC for traceability, reproducibility, and team collaboration.
- Automated workflows with GitHub Actions and deployed model artifacts to staging via Docker containers.
- Integrated Evidently AI to trigger drift alerts and generate summary dashboards for model lifecycle health.

COURSEWORK

Natural Language Processing Specialization – Coursera

Machine Learning Specialization - Coursera

Advance Google Data Analytics Specialization – Coursera

Google Data Analytics Specialization - Coursera

CERTIFICATIONS

Introduction To Python, Data Visualization with Python, Data Analysis with Python – **IBM Carrer**

Data Analytics Job Simulation, PowerBi Job Simulation - **Forage**